

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum  
Internationales Büro



(43) Internationales Veröffentlichungsdatum  
24. Juli 2003 (24.07.2003)

PCT

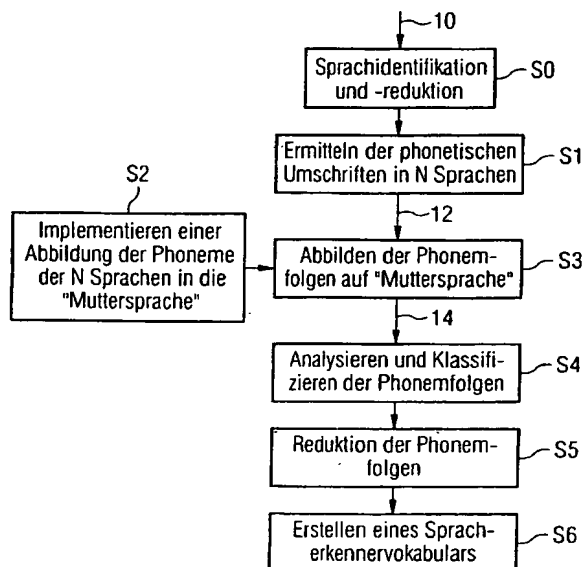
(10) Internationale Veröffentlichungsnummer  
WO 03/060877 A1

- (51) Internationale Patentklassifikation<sup>7</sup>: G10L 15/00, 15/14, 15/16 (71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US): SIEMENS AKTIENGESELLSCHAFT [DE/DE]; Wittelsbacherplatz 2, 80333 München (DE).
- (21) Internationales Aktenzeichen: PCT/EP03/00003 (72) Erfinder; und
- (22) Internationales Anmeldedatum: 2. Januar 2003 (02.01.2003) (75) Erfinder/Anmelder (nur für US): SCHNEIDER, Tobias [DE/DE]; Kranzhornstr. 7, 81825 München (DE).
- (25) Einreichungssprache: Deutsch (74) Gemeinsamer Vertreter: SIEMENS AKTIENGESELLSCHAFT; Postfach 22 16 34, 80506 München (DE).
- (26) Veröffentlichungssprache: Deutsch (81) Bestimmungsstaat (national): US.
- (30) Angaben zur Priorität: 02001256.3 17. Januar 2002 (17.01.2002) EP

[Fortsetzung auf der nächsten Seite]

(54) Title: OPERATING METHOD FOR AN AUTOMATED LANGUAGE RECOGNIZER INTENDED FOR THE SPEAKER-INDEPENDENT LANGUAGE RECOGNITION OF WORDS IN DIFFERENT LANGUAGES AND AUTOMATED LANGUAGE RECOGNIZER

(54) Bezeichnung: BETRIEBSVERFAHREN EINES AUTOMATISCHEN SPRACHERKENNERS ZUR SPRECHERUNABHÄNGIGEN SPRACHERKENNUNG VON WORTEN AUS VERSCHIEDENEN SPRACHEN UND AUTOMATISCHER SPRACHERKENNER



(57) Abstract: The invention relates to an operating method for an automated language recognizer intended for the speaker-independent language recognition of words (10) from different languages, particularly for recognizing names from different languages. Said method is based on a language defined as the mother tongue and has an input phase for establishing a language recognizer vocabulary.

(57) Zusammenfassung: Die Erfindung betrifft ein Betriebsverfahren eines automatischen Spracherkenners zur sprecherunabhängigen Spracherkennung von Worten (10) aus verschiedenen Sprachen, insbesondere zur Erkennung von Namen aus verschiedenen Sprachen, das von einer als Muttersprache definierten Sprache ausgeht und eine Eingabephase zur Erstellung eines Spracherkennervokabulars aufweist.

S0 LANGUAGE IDENTIFICATION AND REDUCTION  
S1 DETECTION OF THE PHONETIC TRANSCRIPTS IN N LANGUAGES  
S2 REPRODUCTION OF THE PHONEMES OF N LANGUAGES IN THE "MOTHER TONGUE"  
S3 REPRODUCTION OF THE PHONEME SEQUENCES IN "MOTHER TONGUE"  
S4 ANALYSIS AND CLASSIFICATION OF THE PHONEME SEQUENCES  
S5 REDUCTION OF THE PHONEME SEQUENCES  
S6 ESTABLISHMENT OF A LANGUAGE RECOGNIZER VOCABULARY

WO 03/060877 A1



(84) **Bestimmungsstaaten (regional):** europäisches Patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR).

**Erklärungen gemäß Regel 4.17:**

- hinsichtlich der Berechtigung des Anmelders, ein Patent zu beantragen und zu erhalten (Regel 4.17 Ziffer ii) für die folgenden Bestimmungsstaaten europäisches Patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR)
- Erfindererklärung (Regel 4.17 Ziffer iv) nur für US

**Veröffentlicht:**

- mit internationalem Recherchenbericht
- vor Ablauf der für Änderungen der Ansprüche geltenden Frist; Veröffentlichung wird wiederholt, falls Änderungen eintreffen

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

DT04 Rec'd PCT/PTO 1 6 JUL 2004

## Beschreibung

Betriebsverfahren eines automatischen Spracherkenners zur sprecherunabhängigen Spracherkennung von Worten aus verschiedenen Sprachen und automatischer Spracherkenner

Die Erfindung betrifft ein Betriebsverfahren eines automatischen Spracherkenners zur sprecherunabhängigen Spracherkennung von Worten aus verschiedenen Sprachen gemäß Patentanspruch 1 und einen entsprechenden automatischen Spracherkenner gemäß Patentanspruch 6.

Für die phonembasierte Spracherkennung ist ein Spracherkennungs-Vokabular erforderlich, das die phonetischen Beschreibungen aller zu erkennender Wörter umfaßt. Dies ist eine Grundvoraussetzung für die phonembasierte Spracherkennung. Wörter werden hierbei durch Phonemfolgen oder -ketten im Vokabular repräsentiert. Während eines Spracherkennungs-Vorgangs wird eine Suche nach dem besten Pfad durch die Phonemfolgen im Vokabular durchgeführt. Diese Suche kann beispielsweise mit dem sogenannten Viterbi-Algorithmus erfolgen. Bei kontinuierlicher Spracherkennung können zudem die Wahrscheinlichkeiten für Übergänge zwischen Wörtern modelliert und in den Viterbi-Algorithmus einbezogen werden.

Die phonetischen Umschriften für die zu erkennenden Wörter sind die Basis der phonembasierten Spracherkennung. Daher stellt sich zu Beginn des Einsatzes eines phonembasierten Spracherkenners immer die Frage, wie derartige phonetische Umschriften gewonnen werden können. Unter phonetischen Umschriften werden hier die phonetischen Beschreibungen der Wörter aus einem Zielvokabular verstanden. Insbesondere stellt sich diese Frage bei Wörtern, die dem Spracherkenner nicht bekannt sind.

Bekannt sind Mobil- oder Schnurlostelefone, die eine sprecherabhängige Namenswahl ermöglichen. Ein Benutzer eines der-

artigen Telefons muß hierzu die im elektronischen Telefonbuch des Telefons enthaltenen Einträge trainieren, um diese später zur Namenswahl per Sprache nutzen zu können. Allerdings kann in der Regel kein anderer Benutzer dieses Feature nutzen, da  
5 die sprecherabhängige Namenswahl nur für eine Person geeignet ist, nämlich für diejenige, welche die Sprachwahl trainiert hat. Um dieses Problem zu umgehen, können die Einträge im elektronischen Telefonbuch in phonetische Umschriften umgewandelt werden.

10 Zum Ermitteln der phonetischen Umschrift aus einem geschriebenen Wort, beispielsweise einem Telefonbucheintrag, sind unterschiedliche Ansätze bekannt. Es sei hier beispielsweise auf die sogenannten Diktiersysteme, die im allgemeinen auf  
15 einem PC zur Ausführung kommen, verwiesen. Bei derartigen Diktiersystemen ist im Normalfall ein Lexikon von typischerweise mehreren 10000 Wörtern mit den Zuordnungen von Buchstabenfolgen zu Phonemfolgen hinterlegt. Da ein solches Lexikon allerdings einen sehr hohen Speicherplatzbedarf aufweist, ist  
20 es für mobile Endgeräte wie beispielsweise Mobil- oder Schnurlostelefone nicht praktikabel.

Bekannt sind auch Systeme, bei denen die Umsetzung eines Wortes in dessen phonetische Umschrift regelbasiert oder durch  
25 speziell trainierte neuronale Netze erfolgt. Diese Verfahren besitzen wie das Lexikon den Nachteil, daß festgelegt werden muß, in welcher Sprache die Phonemfolge realisiert werden soll. Allerdings können insbesondere in elektronischen Telefonbüchern Namen aus verschiedenen Sprachen vorhanden sein.  
30 Eine Umsetzung wäre dann mit dem oben beschriebenen Verfahren nicht oder nur unvollständig möglich.

Daher wurden sogenannte multilinguale Systeme zur Phonemkettenermittlung und Spracherkennung entworfen. Diese Systeme  
35 erlauben die Erzeugung von Phonemketten aus verschiedenen Sprachen.

Schließlich existiert noch eine andere Lösung: Ein Benutzer spricht die Worte in ein Spracherkennungssystem ein, das daraus automatisch Phonemfolgen generiert. Bei großen Wortschätzen, aber auch schon bei einigen dutzend Wörtern, wie beispielsweise bei einem elektronischen Telefonbuch mit 80 Einträgen, ist dies für den Benutzer nicht mehr akzeptabel.

Aufgabe der vorliegenden Erfindung ist es daher, ein Betriebsverfahrens eines automatischen Spracherkenners zur sprecherunabhängigen Spracherkennung von Worten aus verschiedenen Sprachen sowie einen entsprechenden automatischen Spracherkenner vorzuschlagen, welche einfach zu implementieren sind, sich insbesondere zum Einsatz in mobilen Endgeräten eignen, und kostengünstig zu realisieren sind.

Diese Aufgabe wird durch ein Betriebsverfahren mit den Merkmalen des Patentanspruchs 1 und durch einen automatischen Spracherkenner mit den Merkmalen des Patentanspruchs 6 gelöst.

Der Erfindung liegt im wesentlichen die Idee zugrunde, phonetische Umschriften von Wörtern jeweils für N-verschiedene Sprachen zu ermitteln, diese anschließend nachzuverarbeiten und einer phonembasierten einsprachigen Spracherkennung zuzuführen. Diese Vorgehensweise beruht im wesentlichen auf der Erkenntnis, daß ein Benutzer der Spracherkennung normalerweise in seiner Muttersprache spricht. Auch fremdsprachige Wörter, beispielsweise Namen, spricht er normalerweise mit einer "Muttersprachenfärbung", also einem Akzent aus, die bzw. der durch einen sogenannten Muttersprachen-Spracherkenner grob modelliert werden kann. Das Betriebsverfahren geht daher von einer als Muttersprache definierten Sprache aus.

Jede Sprache läßt sich nun mit unterschiedlichen, der jeweiligen Sprache eigenen Phonemen beschreiben. Bekanntermaßen ähneln sich jedoch viele Phoneme verschiedener Sprachen. Ein Beispiel hierfür ist das "p" im englischen und deutschen.

Diese Tatsache wird bei der multilingualen Spracherkennung ausgenutzt. Für ein Ensemble von Sprachen wird hier ein einziges Hidden-Markov-Modell erstellt, mit dem simultan mehrere Sprachen erkannt werden können. Allerdings führt dies zu einem sehr großen Hidden-Markov-Modell, das eine niedrigere Erkennungsrate als ein einsprachiges Hidden-Markov-Modell besitzt. Zudem muß bei einer Erweiterung des Ensembles von Sprachen um beispielsweise eine weitere Sprache ein neues Hidden-Markov-Modell erstellt werden, was sehr aufwendig ist. Dies wird mit der Erfindung vermieden.

Erfindungsgemäß werden in einem ersten Schritt der Eingabephase zur Erstellung eines Spracherkenner-Vokabulars eines Betriebsverfahrens eines automatischen Spracherkenners zur sprecherunabhängigen Spracherkennung von Worten aus verschiedenen Sprachen, insbesondere von Erkennung von Namen aus verschiedenen Sprachen, die phonetischen Umschriften von Wörtern jeweils für N-verschiedene Sprachen ermittelt, um pro Wort N-erste Phonemfolgen entsprechend N-ersten Aussprachevarianten zu erhalten. In einem zweiten Schritt werden die Ähnlichkeiten zwischen den Sprachen ausgenutzt. Hierzu wird eine Abbildung der Phoneme jeder Sprache auf den jeweiligen Phonemsatz der Muttersprache implementiert. Ferner wird in einem dritten Schritt die implementierte Abbildung auf die im ersten Schritt ermittelten N-ersten Phonemfolgen für jedes Wort angewandt. Dadurch werden pro Wort N-zweite Phonemfolgen entsprechend N-zweiten Aussprachevarianten erhalten. Mit dem Muttersprachen-Spracherkenner kann dann bereits eine Anzahl N-verschiedener Sprachen nach Erstellen eines Spracherkenner-Vokabulars mit den im vorhergehenden Schritt erhaltenen N-zweiten Phonemfolgen pro Wort für den Muttersprachen-Spracherkenner erkannt werden.

Die Erfindung hat im wesentlichen die folgenden Vorteile:  
Während ein Look-up-Verfahren in einem Lexikon bei mobilen Endgeräten wegen des großen Speicherplatzbedarfs scheitert und bei der multilingualen Spracherkennung, die für einen

Satz von Sprachen optimiert wurde, für jede neue Sprache neue Hidden-Markov-Modelle erstellt und optimiert werden müssen, wird durch die Grapheme/Phoneme-Konversion in mehrere Sprachen gemäß der Erfindung ein multilinguales System geschaffen, das mit relativ einfachen Mitteln zu implementieren ist, sich daher vor allem zum Einsatz in mobilen Endgeräten eignet und nicht zuletzt kostengünstig zu realisieren ist. Für die Erfindung ist neben der Graphem-zu-Phonem-Umwandlung im wesentlichen nur noch ein Mapping, d. h. ein Abbilden zwischen den einzelnen Sprachen - wie oben erläutert - erforderlich. Die Phonemfolgen-Ermittlung und das anschließende Mapping bzw. Abbilden laufen normalerweise "offline" auf einem Gerät ab, beispielsweise einem Mobiltelefon, einem Personal Digital Assistant oder Personal Computer mit entsprechender Software, und sind daher zeitunkritisch. Die hierfür benötigten Ressourcen können in einem langsamen externen Speicher untergebracht werden.

Da das mit dem oben beschriebenen Verfahren erstellte Spracherkenner-Vokabular jedoch für jedes Wort N-Aussprachevarianten umfaßt, ist der Suchaufwand bei der Spracherkennung groß. Um ihn zu verringern, kann ein weiterer Schritt in das Verfahren eingeführt werden, der noch vor dem Erstellen des Spracherkenner-Vokabulars und nach dem Erzeugen der N-zweiten Phonemfolgen pro Wort ausgeführt wird. In diesem Schritt werden die N-zweiten Phonemfolgen entsprechend den N-zweiten Aussprachevarianten jedes Wort bearbeitet, indem jede zweite Phonemfolge mittels geeigneter Distanzen, insbesondere der Levenshtein-Distanz, analysiert und klassifiziert wird, und die N-zweiten Phonemfolgen jedes Wortes auf wenige, vorzugsweise zwei bis drei, Phonemfolgen reduziert werden, insbesondere indem die Aussprachevarianten weggelassen werden, die der Aussprachevariante der Muttersprache am wenigsten ähnlich sind. Vereinfacht ausgedrückt werden durch diese Reduzierung die weniger wichtigen Aussprachevarianten weggelassen, wodurch sich der Suchaufwand bei der Spracherkennung verringert.

Eine weitere Aufwandsreduktion läßt sich erreichen, indem vor dem ersten Schritt eine Sprachidentifikation und -reduktion vorgenommen wird. Im Rahmen dieser Sprachidentifikation wird für jedes zu erkennende Wort die Wahrscheinlichkeit der Zugehörigkeit zu jeder der N-verschiedenen Sprachen bestimmt. Anhand des Ergebnisses dieser Sprachidentifikation wird die Anzahl der im ersten Verfahrensschritt zu verarbeitenden Sprachen, vorzugsweise auf zwei bis drei verschiedene Sprachen, reduziert. Diese Sprachreduktion erfolgt vorzugsweise, indem die Sprachen mit der geringsten Wahrscheinlichkeit nicht weiterverarbeitet werden. Für ein bestimmtes Wort kann das Ergebnis der Sprachidentifikation beispielsweise wie folgt lauten: "Deutsch 55%, UK-Englisch 16%, US-Englisch 14%, Schwedisch 3%, ...". Bereits nach diesem Ergebnis kann auf drei verschiedene Sprachen reduziert werden, indem Schwedisch weggelassen, d. h. nicht weiterverarbeitet wird.

Das Ermitteln der phonetischen Umschriften im ersten Verfahrensschritt erfolgt vorzugsweise durch mindestens ein neuronales Netz. Neuronale Netze haben sich zum Ermitteln phonetischer Umschriften aus geschriebenen Worten bewährt, da sie gute Ergebnisse hinsichtlich der Genauigkeit und vor allem Verarbeitungsgeschwindigkeit liefern sowie einfach, insbesondere in Software implementierbar sind.

Als Muttersprachen-Spracherkenner kann insbesondere ein Hidden-Markov-Modell zum Einsatz kommen, das für die als Muttersprache definierte Sprache erstellt worden ist.

Die Erfindung betrifft ferner einen Spracherkenner zur sprecherunabhängigen Spracherkennung von Worten aus verschiedenen Sprachen, insbesondere zur Erkennung von Namen aus verschiedenen Sprachen. Hierbei ist eine der verschiedenen Sprachen als Muttersprache definiert. Der Spracherkenner umfaßt

- einen Muttersprachen-Spracherkenner,



- ein erstes Verarbeitungsmodul zum Ermitteln der phonetischen Umschriften von Wörtern jeweils für N-verschiedene Sprachen, um pro Wort N-erste Phonemfolgen entsprechend N-ersten Aussprachevarianten zu erhalten,
  - 5 - ein zweites Verarbeitungsmodul zum Implementieren einer Abbildung der Phonem jeder Sprache auf dem jeweiligen Phonem-satz der Muttersprache,
  - ein drittes Verarbeitungsmodul zum Anwenden der mit dem zweiten Verarbeitungsmodul implementierten Abbildung auf die
  - 10 mit dem ersten Verarbeitungsmodul ermittelten N-ersten Phonemfolgen für jedes Wort, wodurch pro Wort N-zweite Phonemfolgen entsprechend N-zweiten Aussprachevarianten erhalten werden, die mit dem Muttersprachen-Spracherkenner erkannt werden können und
  - 15 - ein viertes Verarbeitungsmodul zum Erstellen eines Sprach-erkenner-Vokabulars mit den durch das dritte Verarbeitungsmodul erhaltenen N-zweiten Phonemfolgen pro Wort für den Muttersprachen-Spracherkenner.
- 20 In einer bevorzugten Ausführungsform umfaßt der automatische Spracherkenner ein fünftes Verarbeitungsmodul zum Bearbeiten der N-zweiten Phonemfolgen entsprechend den N-zweiten Aussprachevarianten jedes Wortes. Das fünfte Verarbeitungsmodul ist derart ausgebildet, daß jede zweite Phonemfolge mittels
- 25 geeigneter Distanzen, insbesondere der Levenshtein-Distanz, analysiert und klassifiziert wird, und die N-zweiten Phonemfolgen jedes Wortes auf wenige, vorzugsweise zwei bis drei, Phonemfolgen reduziert werden.
- 30 Ferner kann der automatische Spracherkenner einen Sprachidentifikator und einen Sprachreduzierer umfassen. Der Sprachidentifikator ist vor das erste Verarbeitungsmodul geschaltet und bestimmt für jedes zu erkennende Wort die Wahrscheinlichkeit der Zugehörigkeit zu jeder der N-verschiedenen Sprachen.
- 35 Der Sprachreduzierer reduziert die Anzahl der vom ersten Verarbeitungsmodul zu verarbeitenden Sprachen, vorzugsweise auf zwei bis drei verschiedene Sprachen, indem die Sprachen mit

der geringsten Wahrscheinlichkeit nicht weiterverarbeitet werden. Sprachidentifikator und Sprachreduzierer verringern sowohl den Verarbeitungsaufwand des automatischen Spracherkenners sowohl in der Eingabephase als auch in der Erkennungsphase beträchtlich.

Vorzugsweise weist das erste Verarbeitungsmodul mindestens ein neuronales Netz zum Ermitteln der phonetischen Umschriften auf.

Schließlich weist der Muttersprachen-Spracherkenner in einer bevorzugten Ausführungsform ein Hidden-Markov-Modell auf, das für die als Muttersprache definierte Sprache erstellt worden ist.

Vorteile und Zweckmäßigkeiten der Erfindung ergeben sich im übrigen aus der nachfolgenden Beschreibung eines Ausführungsbeispiels der Erfindung anhand der einzigen Figur. Diese zeigt ein schematisches Ablaufdiagramm der Eingabephase zur Erstellung eines Spracherkenner-Vokabulars gemäß der Erfindung.

Es soll die sprecherabhängige Namenswahl auf einem Mobiltelefon mit den Namen aus dem Telefonbuch für einen deutschsprachigen Benutzer realisiert werden. In dem Telefonbuch befinden sich neben überwiegend deutschsprachigen Namen auch einige fremdsprachige Namen. Ein Umsetzer für die graphemische Darstellung der Namen ist auf die Sprachen Deutsch, Italienisch, Tschechisch, Griechisch, Türkisch eingestellt, insgesamt als  $N = 5$  verschiedene Sprachen.

In einem Anfangsschritt S0 wird eine Sprachidentifikation der zugeführten Worte 10 bzw. Einträge des Telefonbuchs vorgenommen. Genauer gesagt wird jedes einzelne Wort auf die Wahrscheinlichkeit der Zugehörigkeit zu einer der fünf Sprachen analysiert. Wird beispielsweise ein deutscher Name verarbeitet, so wird die Wahrscheinlichkeit für Deutsch sehr hoch

sein, für die anderen vier Sprachen, nämlich Italienisch, Tschechisch, Griechisch und Türkisch dagegen sehr viel niedriger. Anhand der pro Wort ermittelten Wahrscheinlichkeiten wird die Sprache mit der geringsten Wahrscheinlichkeit für die weitere Verarbeitung weggelassen. Das bedeutet, daß im nachfolgenden Verarbeitungsgang nur noch vier, anstatt fünf Sprachen verarbeitet werden müssen.

In einem ersten Verfahrensschritt S1 wird für jedes Wort die phonetische Umschrift für jede der vier verschiedenen Sprachen ermittelt. Dadurch werden für jedes Wort vier Phonemfolgen entsprechend vier ersten Aussprachevarianten erhalten.

In einem zweiten Verfahrensschritt S2 wird anschließend eine Abbildung der Phoneme jeder der vier Sprachen auf den jeweiligen Phonemsatz der Muttersprache implementiert.

Diese Abbildung wird in einem dritten Verfahrensschritt S3 auf die im ersten Verfahrensschritt S1 erhaltenen vier ersten Phonemfolgen 12 angewandt. Hierdurch werden für jedes Wort vier zweite Phonemfolgen 14 entsprechend vier zweiten Aussprachevarianten erhalten. Die vier zweiten Phonemfolgen 14 können bereits mit einem Muttersprachen-Spracherkenner erkannt werden.

Um allerdings den Verarbeitungsaufwand für den Spracherkenner weiter zu reduzieren, wird pro Wort jede zweite Phonemfolge mittels der Levenshtein-Distanz analysiert und klassifiziert (Schritt S4). Anschließend folgt ein fünfter Verfahrensschritt S5, in dem die analysierten und klassifizierten zweiten Phonemfolgen pro Wort auf drei Phonemfolgen reduziert werden.

Schließlich wird in einem letzten Schritt S6 ein Spracherkenner-Vokabular mit dem in dem fünften Verfahrensschritt S5 erhaltenen drei zweiten Phonemfolgen pro Wort für den Muttersprachen-Spracherkenner erstellt. Durch die nochmalige Reduk-

tion der Phonemfolgen im fünften Verfahrensschritt S5 wird also das zu speichernde und während einer Spracherkennung zu durchsuchende Spracherkenner-Vokabular merklich reduziert. Dies bringt in einer praktischen Anwendung der Spracherken-  
5 nung einerseits den Vorteil eines geringeren Speicherplatzbedarfs und andererseits einer schnelleren Verarbeitung, da ein kleineres Vokabular durchsucht werden muß.

Nach Ablauf des beschriebenen Verfahrens kann der Benutzer  
10 mittels Spracherkennung eine Namenswahl, also den sprachgesteuerten Aufruf gespeicherter Rufnummern über den Namen des Teilnehmers vornehmen, ohne daß er den Namen des zu rufenden Teilnehmers explizit einmal vorsprechen, also trainieren, muß.

15 Im folgenden wird kurz erläutert, was der Benutzer des Mobiltelefons zur Verbesserung der Spracherkennung tun kann. Sollte er einmal feststellen, daß ein bestimmter Name nicht gut erkannt wird, kann er das Spracherkenner-Menü seines Mobilte-  
20 lefons aufrufen und dort die Anwendung "Namenswahl" auswählen. Unter dieser Anwendung kann ihm nun eine oder auch mehrere Möglichkeiten angeboten werden, um die Spracherkennung eines bestimmten Wortes, genauer gesagt eines bestimmten Namens aus dem elektronischen Telefonbuch des Mobiltelefons zu  
25 verbessern. Im folgenden werden beispielhaft einige dieser Möglichkeiten kurz erläutert:

1. Der Benutzer kann das schlecht oder gar nicht erkannte Wort nochmals in das Mobiltelefon einsprechen und anschlie-  
30 ßend durch den im Mobiltelefon enthaltenen Spracherkenner in eine Phonemfolge umsetzen lassen. In diesem Fall werden vorher automatisch ermittelte Aussprachevarianten ganz oder teilweise, je nachdem welche Nähe sie zu der neu ermittelten Phonemfolge besitzen, aus dem Vokabular des Spracherkenners  
35 entfernt.

2. Alternativ kann sich der Benutzer auf dem Display des Mobiltelefons eine Art Lautschrift des schlecht oder gar nicht erkannten Eintrags des elektronischen Telefonbuchs anzeigen lassen. Die Art Lautschrift kann der Benutzer dann bei Nicht-  
5 zutreffen, d. h. bei schlechter Übereinstimmung mit seiner Aussprache, editieren. Beispielsweise kann durch die automatische Umsetzung des Eintrags "Jacques Chirac" als Lautschrift "Jakwes Schirack" gespeichert sein. Erscheint nun dem Benutzer diese Lautschrift als fehlerhaft, kann er sie mit-  
10 tels seines Mobiltelefons editieren, beispielsweise zu "Schack Schirack". Anschließend kann das System dazu die phonetische Beschreibung ermittelt und diese ins Spracherkenner-Vokabular neu aufnehmen. Damit sollte die automatische Spracherkennung zuverlässig funktionieren.

15 3. Schließlich kann der Benutzer durch eine explizite Angabe der Sprache, aus der ein fehlerhaft oder gar nicht erkannter Name stammt oder durch explizite Auswahl einer bestimmten Sprache für einen bestimmten Namen die Erkennung wesentlich  
20 verbessern. In einem derartigen Fall werden alle Aussprachevarianten für den Namen aus dem Spracherkenner-Vokabular entfernt, die nicht der explizit angegebenen Sprache zugeordnet sind.

25 Die Erfindung kann auch vorteilhaft in anderen mobilen Geräten außer einem Mobiltelefon z. B. einem Personal Assistant oder auch einem Personal Computer verwendet, d. h. eingesetzt werden.

## Patentansprüche

1. Betriebsverfahren eines automatischen Spracherkenners zur sprecherunabhängigen Spracherkennung von Worten (10) aus verschiedenen Sprachen, insbesondere zur Erkennung von Namen aus verschiedenen Sprachen, das von einer als Muttersprache definierten Sprache ausgeht und eine Eingabephase zur Erstellung eines Spracherkenner-Vokabulars mit den folgenden Schritte aufweist:
- 5 (a) Ermitteln der phonetischen Umschriften von Wörtern jeweils für N verschiedene Sprachen, um pro Wort N erste Phonemfolgen (12) entsprechend N ersten Aussprachevarianten zu erhalten (S1),
- (b) Implementieren einer Abbildung der Phoneme jeder Sprache auf den jeweiligen Phonemsatz der Muttersprache (S2),
- 15 (c1) Anwenden der in Schritt (b) implementierten Abbildung auf die in Schritt (a) ermittelten N ersten Phonemfolgen (12) für jedes Wort, wodurch pro Wort N zweite Phonemfolgen (14) entsprechend N zweiten Aussprachevarianten erhalten werden, die mit einem Muttersprachen-Spracherkenner erkannt werden können (S3), und
- 20 (d) Erstellen eines Spracherkenner-Vokabulars mit den im vorhergehenden Schritt erhaltenen N zweiten Phonemfolgen pro Wort für den Muttersprachen-Spracherkenner (S6).
- 25
2. Betriebsverfahren nach Anspruch 1, ferner gekennzeichnet durch die folgenden, vor dem Schritt (d) und nach dem Schritt (c1) auszuführenden Schritte:
- (c2) Bearbeiten der N zweiten Phonemfolgen (14) entsprechend den N zweiten Aussprachevarianten jedes Wortes, indem
- 30 (c21) jede zweite Phonemfolge (14) mittels geeigneter Distanzen, insbesondere der Levenshtein-Distanz, analysiert und klassifiziert wird (S4), und
- (c22) die N zweiten Phonemfolgen jedes Wortes auf wenige,
- 35 vorzugsweise 2 bis 3, Phonemfolgen reduziert werden (S5).

3. Betriebsverfahren nach Anspruch 1 oder 2,  
dadurch gekennzeichnet, dass  
vor dem Schritt (a) eine Sprachidentifikation vorgenommen  
wird, durch die für jedes zu erkennende Wort die Wahr-  
5      scheinlichkeit der Zugehörigkeit zu jeder der N verschiedenen Spra-  
chen bestimmt wird, und anhand des Ergebnisses der Sprach-  
identifikation die Anzahl der in Schritt (a) zu verarbeiten-  
den Sprachen, vorzugsweise auf 2 bis 3 verschiedene Sprachen,  
reduziert werden, indem die Sprachen mit der geringsten Wahr-  
10    scheinlichkeit nicht weiter verarbeitet werden (S0).

4. Betriebsverfahren nach einem der Ansprüche 1 bis 3,  
dadurch gekennzeichnet, dass  
das Ermitteln der phonetischen Umschriften in Schritt (a)  
15    durch mindestens ein neuronales Netz erfolgt.

5. Betriebsverfahren nach einem der Ansprüche 1 bis 4,  
dadurch gekennzeichnet, dass  
als Muttersprachen-Spracherkenner ein Hidden-Markov-Modell  
20    eingesetzt wird, das für die als Muttersprache definierte  
Sprache erstellt worden ist.

6. Automatischer Spracherkenner zur sprecherunabhängigen  
Spracherkennung von Worten aus verschiedenen Sprachen, insbe-  
25    sondere zur Erkennung von Namen aus verschiedenen Sprachen,  
wobei eine der verschiedenen Sprachen als Muttersprache defi-  
niert ist, mit  
- einem Muttersprachen-Spracherkenner,  
- einem ersten Verarbeitungsmodul zum Ermitteln der phoneti-  
30    schen Umschriften von Wörtern jeweils für N verschiedene  
Sprachen, um pro Wort N erste Phonemfolgen entsprechend N  
ersten Aussprachevarianten zu erhalten,  
- einem zweiten Verarbeitungsmodul zum Implementieren einer  
Abbildung der Phoneme jeder Sprache auf den jeweiligen Pho-  
35    nemsatz der Muttersprache,  
- einem dritten Verarbeitungsmodul zum Anwenden der mit dem  
zweiten Verarbeitungsmodul implementierten Abbildung auf die

mit dem ersten Verarbeitungsmodul ermittelten N ersten Phonemfolgen für jedes Wort, wodurch pro Wort N zweite Phonemfolgen entsprechend N zweiten Aussprachevarianten erhalten werden, die mit dem Muttersprachen-Spracherkenner erkannt

5 werden können, und

- einem vierten Verarbeitungsmodul zum Erstellen eines Spracherkenner-Vokabulars mit den durch das dritte Verarbeitungsmodul erhaltenen N zweiten Phonemfolgen pro Wort für den Muttersprachen-Spracherkenner.

10

7. Automatischer Spracherkenner nach Anspruch 6, gekennzeichnet durch

ein fünftes Verarbeitungsmodul zum Bearbeiten der N zweiten Phonemfolgen entsprechend den N zweiten Aussprachevarianten

15 jedes Wortes, das derart ausgebildet ist, dass jede zweite Phonemfolge mittels geeigneter Distanzen, insbesondere der Levenshtein-Distanz, analysiert und klassifiziert wird, und die N zweiten Phonemfolgen jedes Wortes auf wenige, vorzugsweise 2 bis 3, Phonemfolgen reduziert werden.

20

8. Automatischer Spracherkenner nach Anspruch 6 oder 7, gekennzeichnet durch

einen Sprachidentifikator, der vor das erste Verarbeitungsmodul geschaltet ist und für jedes zu erkennende Wort die Wahr-

25 scheinlichkeit der Zugehörigkeit zu jeder der N verschiedenen Sprachen bestimmt, und einen Sprachreduzierer, der die Anzahl

der vom ersten Verarbeitungsmodul zu verarbeitenden Sprachen, vorzugsweise auf 2 bis 3 verschiedene Sprachen, reduziert

30 werden, indem die Sprachen mit der geringsten Wahrscheinlichkeit nicht weiter verarbeitet werden.

9. Automatischer Spracherkenner nach einem der Ansprüche 6 bis 8,

dadurch gekennzeichnet, dass

35 das erste Verarbeitungsmodul mindestens ein neuronales Netz zum Ermitteln der phonetischen Umschriften aufweist.



10. Automatischer Spracherkenner nach einem der Ansprüche 6 bis 9,

dadurch gekennzeichnet, dass

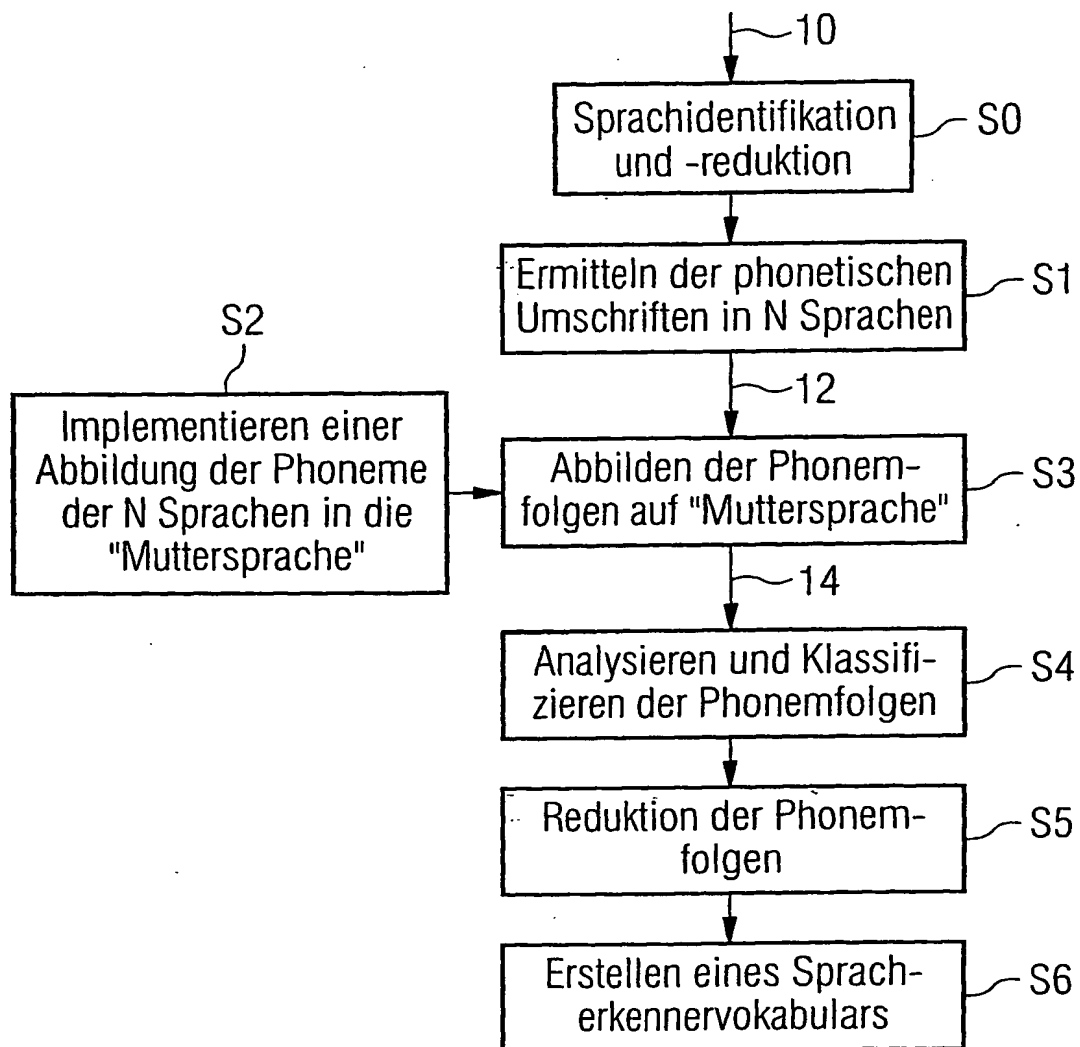
der Muttersprachen-Spracherkenner ein Hidden-Markov-Modell

5 aufweist, das für die als Muttersprache definierte Sprache erstellt worden ist.

11. Verwendung des Betriebsverfahrens nach einem der Ansprüche 1 bis 5 und des automatischen Spracherkenners nach einem  
10 der Ansprüche 6 bis 10 in einem mobilen Gerät wie einem Mobiltelefon, Personal Digital Assistant oder einem Personal Computer.

**THIS PAGE BLANK (USPTO)**

1/1



DT04 Rec'd PCT 1 6 JUL 2004

**THIS PAGE BLANK (USPTO)**

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/EP 03/00003

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 G10L15/00 G10L15/14 G10L15/16

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, INSPEC

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 95 02879 A (BORDEAUX THEODORE AUSTIN) 26 January 1995 (1995-01-26) abstract; figure 9 page 5, line 28 -page 7, line 12 ---	1,4-6,9, 10
A	US 6 085 160 A (D'HOORE ET AL) 4 July 2000 (2000-07-04) abstract column 1, line 61 -column 2, line 13 ---	1,6,11
A	US 5 805 771 A (NETSCH LORIN P ET AL) 8 September 1998 (1998-09-08) abstract; figures 3,4 ---	3-5,8-10
	--- -/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \* & \* document member of the same patent family

Date of the actual completion of the international search

13 May 2003

Date of mailing of the international search report

28/05/2003

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Quélavoine, R

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/EP 03/00003

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MUTHUSAMY Y K ET AL: "AUTOMATIC SEGMENTATION AND IDENTIFICATION OF TEN LANGUAGES USING TELEPHONE SPEECH" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (ICSLP). BANFF, vol. 2, 12 - 16 October 1992, pages 1007-1010, XP000871623 EDMONTON, UNIVERSITY OF ALBERTA, CA abstract	3,4,8,9
A	VAICH T ET AL: "HMM phoneme recognition with supervised training and Viterbi algorithm" EIGHTEENTH CONVENTION OF ELECTRICAL AND ELECTRONICS ENGINEERS IN ISRAEL (CAT. NO.95TH8044), TEL AVIV, ISRAEL, 7 - 8 March 1995, pages 3.2.1/1-5, XP002198892 1995, New York, NY, USA, IEEE, USA ISBN: 0-7803-2498-6 abstract	2,7
A	US 6 212 500 B1 (KÖHLER) 3 April 2001 (2001-04-03) abstract	1,6,11
A	ANDERSON O ET AL: "THE ONOMASTICA INTERLANGUAGE PRONUNCIATION LEXICON" 4TH EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY. EUROSPEECH '95. MADRID, SPAIN, vol. 1 CONF. 4, 18 - 21 September 1995, pages 829-832, XP000854833 MADRID: GRAFICAS BRENS, ES abstract	1,6,11

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 03/00003

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9502879	A	26-01-1995	AT 200590 T 15-04-2001
		AU 682380 B2 02-10-1997	
		AU 7328294 A 13-02-1995	
		CA 2167200 A1 26-01-1995	
		DE 69427083 D1 17-05-2001	
		DE 69427083 T2 06-12-2001	
		EP 0708958 A1 01-05-1996	
		JP 9500223 T 07-01-1997	
		WO 9502879 A1 26-01-1995	
		US 5758023 A 26-05-1998	
US 6085160	A	04-07-2000	EP 1095371 A1 02-05-2001
		WO 0003386 A1 20-01-2000	
		JP 2002520664 T 09-07-2002	
US 5805771	A	08-09-1998	NONE
US 6212500	B1	03-04-2001	DE 19636739 C1 03-07-1997
		AT 209814 T 15-12-2001	
		BR 9712979 A 31-10-2000	
		CN 1230277 A 29-09-1999	
		CN 1237259 A 01-12-1999	
		WO 9811537 A2 19-03-1998	
		WO 9811534 A1 19-03-1998	
		DE 59705581 D1 10-01-2002	
		EP 0925461 A2 30-06-1999	
		EP 0925579 A1 30-06-1999	
		ES 2169432 T3 01-07-2002	
		JP 2000505914 T 16-05-2000	
		JP 2001503154 T 06-03-2001	
		US 6460017 B1 01-10-2002	

**THIS PAGE BLANK (USPTO)**



# INTERNATIONALES RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/EP 03/00003

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES  
IPK 7 G10L15/00 G10L15/14 G10L15/16

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

## B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)  
IPK 7 G10L

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal, WPI Data, INSPEC

## C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	WO 95 02879 A (BORDEAUX THEODORE AUSTIN) 26. Januar 1995 (1995-01-26) Zusammenfassung; Abbildung 9 Seite 5, Zeile 28 -Seite 7, Zeile 12 ----	1,4-6,9, 10
A	US 6 085 160 A (D'HOORE ET AL) 4. Juli 2000 (2000-07-04) Zusammenfassung Spalte 1, Zeile 61 -Spalte 2, Zeile 13 ----	1,6,11
A	US 5 805 771 A (NETSCH LORIN P ET AL) 8. September 1998 (1998-09-08) Zusammenfassung; Abbildungen 3,4 ----- -/--	3-5,8-10

☒ Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

☒ Siehe Anhang Patentfamilie

\* Besondere Kategorien von angegebenen Veröffentlichungen :

\*A\* Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

\*E\* älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

\*L\* Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

\*O\* Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

\*P\* Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

\*T\* Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

\*X\* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

\*Y\* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

\*Z\* Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

13. Mai 2003

Absendedatum des internationalen Recherchenberichts

28/05/2003

Name und Postanschrift der internationalen Recherchenbehörde  
Europäisches Patentamt, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Quélavoine, R

# INTERNATIONALE RECHERCHENBERICHT

Internationales Aktenzeichen  
PCT/EP 03/00003

## C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	MUTHUSAMY Y K ET AL: "AUTOMATIC SEGMENTATION AND IDENTIFICATION OF TEN LANGUAGES USING TELEPHONE SPEECH" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (ICSLP). BANFF, Bd. 2, 12. - 16. Oktober 1992, Seiten 1007-1010, XP000871623 EDMONTON, UNIVERSITY OF ALBERTA, CA Zusammenfassung	3,4,8,9
A	VAICH T ET AL: "HMM phoneme recognition with supervised training and Viterbi algorithm" EIGHTEENTH CONVENTION OF ELECTRICAL AND ELECTRONICS ENGINEERS IN ISRAEL (CAT. NO.95TH8044), TEL AVIV, ISRAEL, 7. - 8. März 1995, Seiten 3.2.1/1-5, XP002198892 1995, New York, NY, USA, IEEE, USA ISBN: 0-7803-2498-6 Zusammenfassung	2,7
A	US 6 212 500 B1 (KÖHLER) 3. April 2001 (2001-04-03) Zusammenfassung	1,6,11
A	ANDERSON O ET AL: "THE ONOMASTICA INTERLANGUAGE PRONUNCIATION LEXICON" 4TH EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY. EUROSPEECH '95. MADRID, SPAIN, Bd. 1 CONF. 4, 18. - 21. September 1995, Seiten 829-832, XP000854833 MADRID: GRAFICAS BRENS, ES Zusammenfassung	1,6,11

# INTERNATIONALE RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/EP 03/00003

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
WO 9502879	A	26-01-1995	AT 200590 T 15-04-2001
			AU 682380 B2 02-10-1997
			AU 7328294 A 13-02-1995
			CA 2167200 A1 26-01-1995
			DE 69427083 D1 17-05-2001
			DE 69427083 T2 06-12-2001
			EP 0708958 A1 01-05-1996
			JP 9500223 T 07-01-1997
			WO 9502879 A1 26-01-1995
			US 5758023 A 26-05-1998
US 6085160	A	04-07-2000	EP 1095371 A1 02-05-2001
			WO 0003386 A1 20-01-2000
			JP 2002520664 T 09-07-2002
US 5805771	A	08-09-1998	KEINE
US 6212500	B1	03-04-2001	DE 19636739 C1 03-07-1997
			AT 209814 T 15-12-2001
			BR 9712979 A 31-10-2000
			CN 1230277 A 29-09-1999
			CN 1237259 A 01-12-1999
			WO 9811537 A2 19-03-1998
			WO 9811534 A1 19-03-1998
			DE 59705581 D1 10-01-2002
			EP 0925461 A2 30-06-1999
			EP 0925579 A1 30-06-1999
			ES 2169432 T3 01-07-2002
			JP 2000505914 T 16-05-2000
			JP 2001503154 T 06-03-2001
			US 6460017 B1 01-10-2002

**THIS PAGE BLANK (USPTO)**